



An Evaluation of NJSO CHAMPS
Findings from the 2013-2014 Season

Steven J. Holochwost

WolfBrown

September 30, 2014

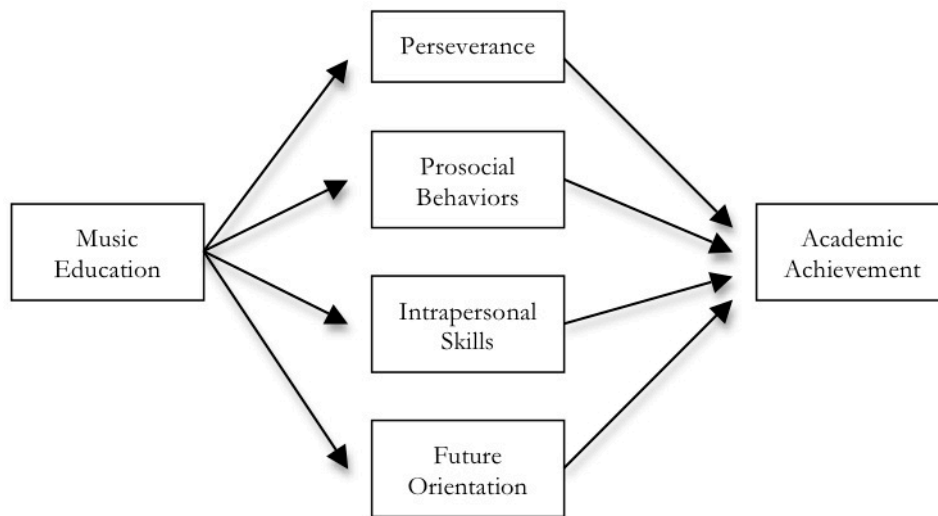
Table of Contents

Executive Summary.....	3
Introduction.....	7
Methods.....	10
Results.....	15
Discussion.....	24
Works Cited.....	27
Appendix A.....	29

Executive Summary

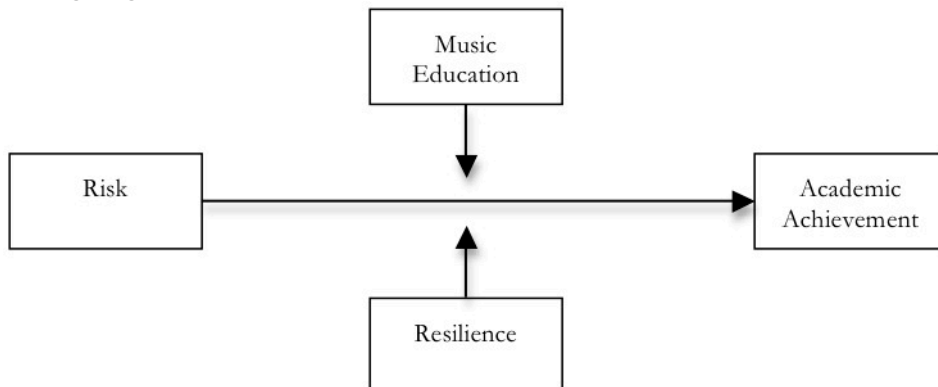
An education in music has been linked to better performance on assessments of reading and math, but it is not known how music education may achieve these effects. It is fair to ask why improved ability in academic domains such as reading and math would result from instruction in music. One possibility is that music education fosters the development of other more proximal domains such as perseverance or prosocial behaviors, and that it is through fostering development in these areas that music education may ultimately influence distal outcomes such as academic achievement. Thus the effects of music education on academic achievement may be transmitted or mediated through the changes this education exerts on these more proximate outcomes. This possibility is summarized in the following diagram.

Figure S1. Mediated Effects of Music Education



Another possibility is that music education buffers or mitigates the effects of environmental risk – poverty and associated factors such as single parenthood and low levels of parental education – on academic achievement or more proximal outcomes. This may be achieved by amplifying the effects of other sources of resilience in a child’s environment, such as the regular observance of family routines or high levels of social support. Thus music education may buffer the effects of risk on academic achievement, either alone or in concert with other sources of resilience, as shown below.

Figure S2. Mitigating Effects of Music Education



These two accounts of how a program like the New Jersey Symphony Orchestra's Character, Achievement, and Music Project (NJSO CHAMPS) may benefit children led to the formation of two sets of questions that guided the evaluation:

Guiding Questions, Set 1: Direct and Mediating Effects

- A. Is experience in CHAMPS associated with higher levels of academic achievement?
- B. Is it associated with higher levels in one or more of the proximate domains of perseverance, prosocial behaviors, intrapersonal skills, or future orientation?
- C. If experience in CHAMPS is associated with academic achievement and at least one proximal outcome, is it the case that the effects of program experience on academic achievement are mediated by its effects of those outcomes?

Guiding Questions, Set 2: Indirect or Mitigating Effects

- A. Does risk predict lower levels of academic achievement or poorer performance on measures of one or more proximal outcomes?
- B. Assuming that it does, are these effects mitigated or buffered by experience in the program?
- C. Are these mitigating effects amplified by other sources of resilience in the child's environment?

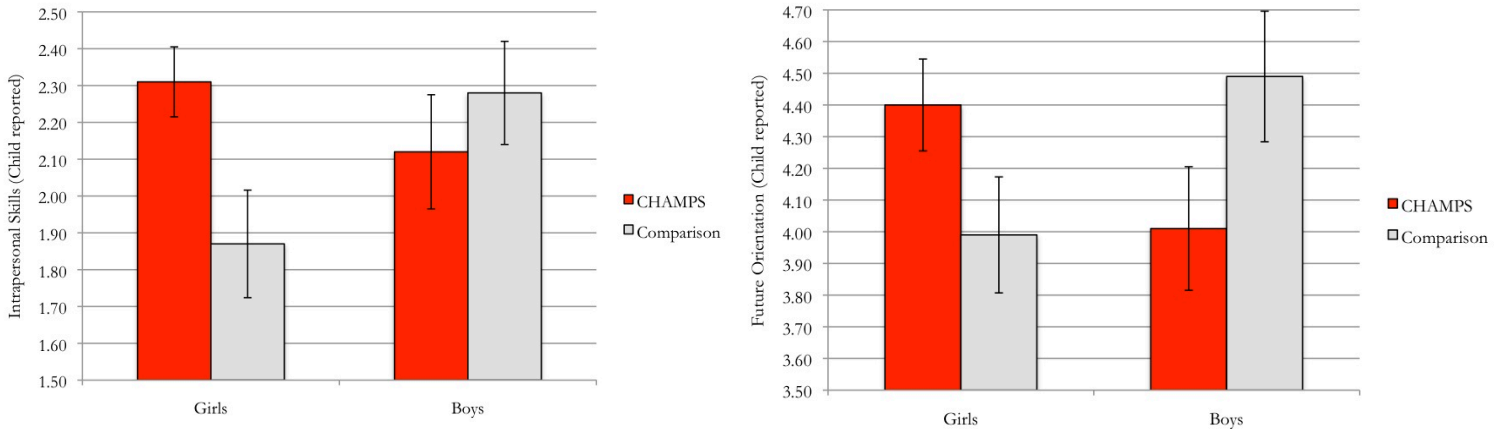
Two other guiding questions do not stem directly from the literature reviewed above, but are nevertheless crucial to assessing the efficacy of CHAMPS. First, it is necessary to establish whether and to what extent students enrolled in the program demonstrate positive change in their musical achievement. Second, it is of interest to examine whether each student's degree of improvement in their musical achievement is predicted by their level of engagement in the program.

To address these questions, data were collected from 36 students enrolled in CHAMPS and a comparison group of 24 students attending the same grades and classrooms at University Heights Charter School (UHCS). A series of measures were collected from students, their parents or caregivers, their UHCS teachers, and their primary instructors in CHAMPS. Most measures were collected twice, coinciding with the first and final two weeks of the CHAMPS program (though see the Methods and Results sections below for exceptions), and these were supplemented by year-end data provided by UHCS.

The guiding questions were addressed using a series of statistical models that controlled for (i.e., held constant) the effects of a child's gender, grade, and classroom on their post-CHAMPS assessments. The analyses also controlled for students' scores on pre-CHAMPS assessments wherever those data were available. Controlling for pre-CHAMPS scores allows us to attribute differences in *post*-CHAMPS scores to the program's influence with more confidence, and constitute a stronger form of evidence of the program's effects.

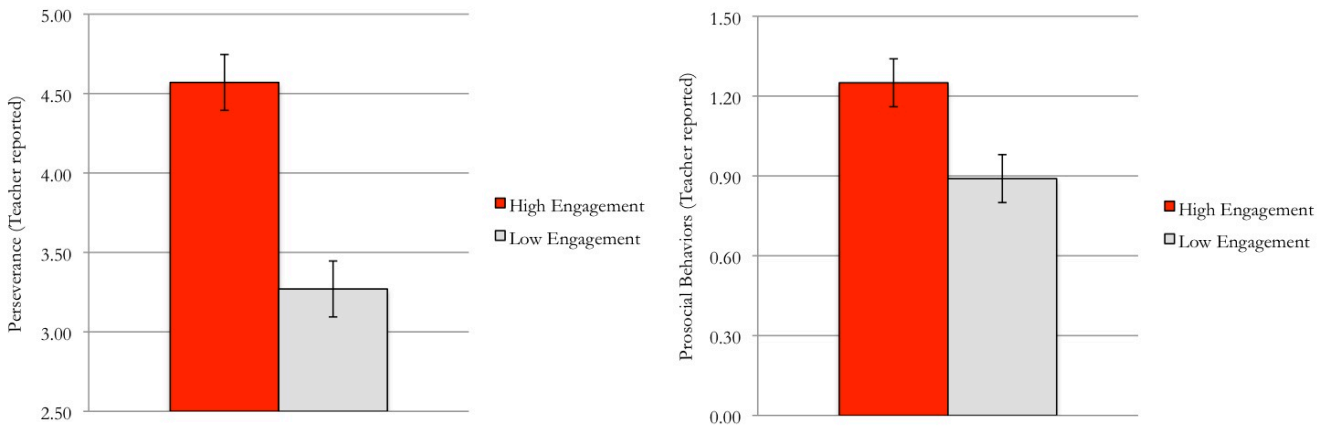
The guiding questions were first examined from the perspective of enrollment: whether enrollment in CHAMPS during the 2013-2014 academic year was associated with higher levels of either academic achievement or more proximal outcomes. While there was no evidence that enrollment in CHAMPS was associated with higher levels of academic achievement, enrollment was associated with higher parent ratings of children's prosocial behaviors, as well as higher child ratings of their own intrapersonal skills and future orientation for girls (see Figure S3). Differences between students enrolled in CHAMPS and their peers were observed for a number of other indicators of prosocial behavior, including absences, the number of scholar dollars, and tardies, though these differences did not achieve statistical significance.

Figure S3. Differences in Child-Reported Intrapersonal Skills and Future Orientation by Enrollment



Next we examined whether a student’s level of engagement in CHAMPS – as rated by their program teachers – would predict differences in either academic achievement or more proximal outcomes. More highly-engaged students of both genders were rated as exhibiting higher levels of perseverance by UHCS teachers (see Figure S4), while parents of more highly-engaged boys also rated their children as exhibiting higher levels of perseverance. Teachers also rated more highly-engaged students as displaying higher levels of prosocial behavior, while more highly-engaged boys reported higher levels of prosocial behavior for themselves. More highly-engaged students were also found to have higher totals of year-end scholar dollars (awarded by UCHS based on behavior).

Figure S4. Differences in Teacher-Reported Perseverance and Prosocial Behaviors by Engagement



In many cases, the sizes of the effects associated with both enrollment and engagement in CHAMPS were large. For example, the effect size of enrollment in CHAMPS on parent-rated prosocial behaviors was 0.75; an effect size of 0.80 or greater is considered large, while an effect between 0.40 and 0.70 is considered medium or moderate. For girls, enrollment in CHAMPS was associated with large effect sizes for ratings of their own intrapersonal skills and future orientation. The finding of a large, positive effect for CHAMPS enrollment on girls' perceptions of their prospects for the future is especially interesting, as early adolescence is a period during which girls' self-esteem declines precipitously. Enrollment in CHAMPS may exert a recalibrating effect, correcting girls' misconceptions about their abilities and prospects for the future (this may also explain why the direction of this effect was in the opposite direction for boys; see the Discussion section of the full report).

Higher levels of engagement in CHAMPS were also associated with benefits to other proximate domains, though these benefits were realized disproportionately by boys. More highly engaged children were rated as being more perseverant by their parents; for girls, the size of this effect was moderate (0.44), but for boys it was very large (14.8). Similarly, children's ratings of their own prosocial behaviors were higher for more highly engaged children, but whereas for girls the size of this effect was modest (0.19), for boys it was large (3.10). These results suggest that for certain proximate outcomes there may be both threshold and dosage effects for CHAMPS that vary as a function of gender. For girls, merely being enrolled in CHAMPS is associated with benefits to intrapersonal strengths and future orientation, while for boys the most substantial benefits were realized among those who were most highly engaged.

The second set of guiding questions examined the possibility that either enrollment or engagement in CHAMPS might exert an indirect effect on academic achievement or proximal outcomes by mitigating or buffering the effects of risk. Thus the first question of this set sought to establish whether risk did in fact exert deleterious effects on various outcomes within the context of the study sample. However, neither risk (defined as described above) nor poverty (defined as income-to-needs ratio) predicted levels of academic achievement or any other measure, with the exception of school absences. Without an effect of risk, there was nothing for enrollment or engagement in CHAMPS to mitigate, nor was there a possibility that these mitigating effects might be amplified by other sources of resilience in the child's home.

The final set of guiding questions concerned students' musical progress, and to what extent this progress was predicted by each student's level of engagement in the program. The mean score on the post-CHAMPS assessment of musical achievement was nearly twice that observed at the pre-CHAMPS assessment, with individual students exhibiting statistically-significant changes in scores and nearly three-quarters of students exhibited positive gains. Higher levels of engagement were associated with higher post-CHAMPS scores after controlling for pre-CHAMPS scores.

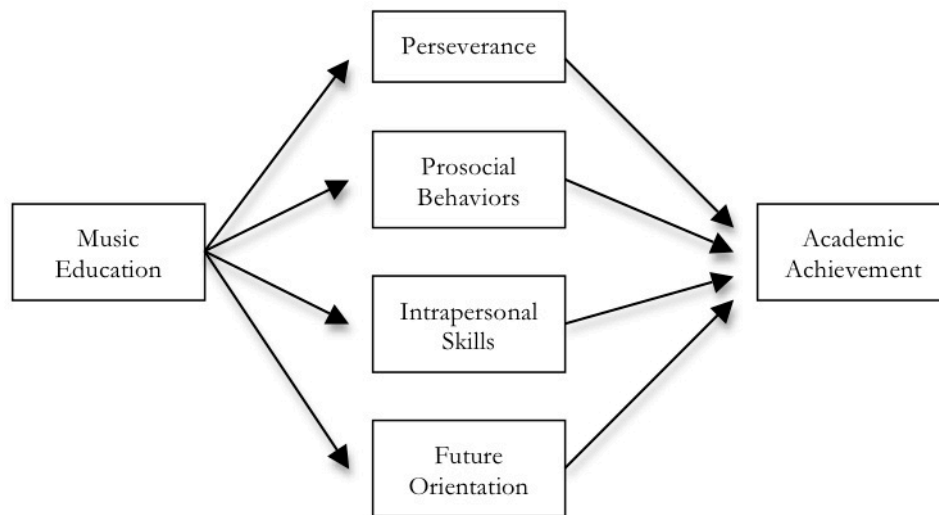
This evaluation was not without limitations. In particular, the small sample size constrained statistical power, rendering the effects of enrollment or engagement in CHAMPS more difficult to detect. Nevertheless, the evaluation did provide evidence that CHAMPS may contribute to the development of key outcomes, including perseverance, prosocial behaviors, intrapersonal skills, and future orientation. Fostering these characteristics in students can have long-lasting consequences for life outcomes that are in the end more important than grades or test scores, and achieving those long-lasting effects is the ultimate goal of CHAMPS: to harness the power of music to change lives.

Introduction

Arts education is associated with better academic outcomes for children. In particular, an education in music has been linked to better performance on assessments of reading and math.¹ However, it is not known how music education may achieve these effects. It is fair to ask why improved ability in academic domains such as reading and math would result from instruction in music.

One possibility is that music education fosters the development of other more proximal domains, and that it is through fostering development in these areas that music education may ultimately influence distal outcomes such as academic achievement. For example, the discipline of learning to play an instrument – through which what once seemed impossible becomes routine through the application of sustained effort – may foster students’ perseverance, defined by Duckworth and her colleagues as “maintaining effort and interest over years despite failure, adversity, and plateaus in progress.”² Students’ ability to relate in positive ways with their peers and to manage their behavior to meet others’ expectations may benefit from the opportunity to play as part of an ensemble under the direction of an instructor. In addition to allowing students to develop these prosocial behaviors or interpersonal skills in this context may offer students unique opportunities to develop their *intrapersonal* skills as well. For example, a student who learns to accept constructive criticism about their playing from peers and teachers may be more willing to accept criticism in other domains. Finally, the opportunity to learn from teachers who have attended college, traveled and performed – as well as the opportunity to performs themselves – may change the ways in which students perceive themselves, their place in the world, and their prospects for the future. Thus the effects of music education on the distal outcome of academic achievement may be transmitted or mediated through the changes this education exerts on these more proximate outcomes. This possibility is summarized in the following diagram.

Figure 1. Mediated Effects of Music Education

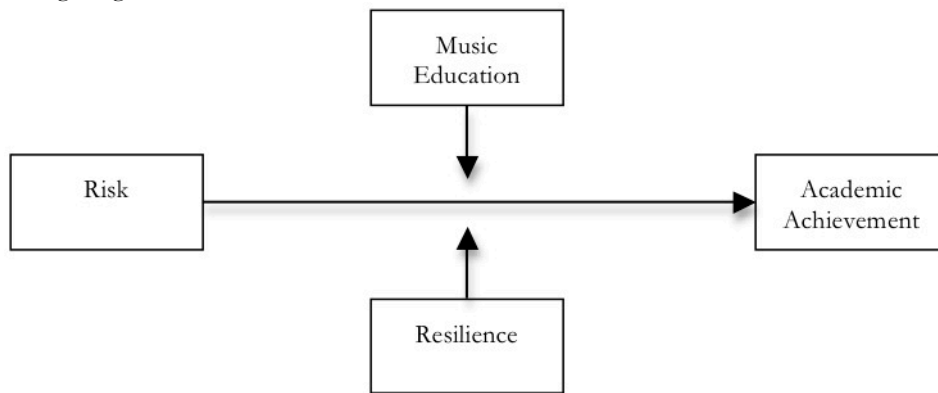


¹ Winner, E., Goldstein, T., & Vincent-Lancrin, S. (2013). *Art for Art's Sake? The Impact of Arts Education*. Washington, DC: Organization for Economic Cooperation and Development.

² Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology*, 92, 1087-1101. Quotation appears on p. 1087.

Another possibility is that music education buffers or mitigates the effects of environmental risk on academic achievement. It has long been known that risk – poverty and associated factors such as single parenthood and low levels of parental education – is associated with lower levels of academic achievement.³ The literature documenting the effects of risk on more proximate domains such as prosocial behaviors is not as large, but certainly there is evidence that risk exerts deleterious effects on these domains as well.⁴ Music education may have the capacity to reduce the impact of risk on academic achievement, proximate domains, or both. This may be achieved by amplifying the effects of other sources of resilience in a child’s environment. The effects of music education may be greater for a child from a household that observes regular family routines (such as sitting down to dinner together) or in which family members express positive feelings toward one another, or where the parent enjoys high levels of social support or an especially close relationship with their child. Thus music education may buffer the effects of risk on academic achievement, either alone or in concert with other sources of resilience, as shown below.

Figure 2. Mitigating Effects of Music Education



Note that the same relationships among risk, music education, and resilience may be observed with respect to any or all of the proximate outcomes discussed above. That is, just as the program may buffer the effects of risk on academic achievement, so too might it buffer (either alone or in combination with other resilience factors) the effects of risk on perseverance, prosocial behaviors, intrapersonal skills, or future orientation.

³ Brooks-Gunn, J. & Duncan, G. J. (1997). The effects of poverty on children. *The Future of Children*, 7, 55-71.

⁴ McLoyd, V. C. (1998). Socioeconomic disadvantage and child development. *American Psychologist*, 53, 185-204.

These two accounts of how a program like the New Jersey Symphony Orchestra's Character, Achievement, and Music Project (CHAMPS) may benefit children led to the formation of two sets of questions that guided the evaluation:

Set 1: Direct and Mediation Effects

- A. Is experience in CHAMPS associated with higher levels of academic achievement?
- B. Is it associated with higher levels in one or more of the proximate domains of perseverance, prosocial behaviors, intrapersonal skills, or future orientation?
- C. If experience in CHAMPS is associated with academic achievement and at least one proximal outcome, is it the case that the effects of program experience on academic achievement are mediated by its effects of those outcomes?

Set 2: Indirect or Mitigating Effects

- A. Does risk predict lower levels of academic achievement or poorer performance on measures of one or more proximal outcomes?
- B. Assuming that it does, are these effects are mitigated or buffered by experience in the program?
- C. Are these mitigating effects amplified by other sources of resilience in the child's environment?

Note that in answering these questions program experience may be defined in two ways: first, as enrollment in CHAMPS, and second, by a given enrollee's level of engagement in the program. It is noted which definition was used and why throughout the results and discussion sections that follow.

Two other guiding questions do not stem directly from the literature reviewed above, but are nevertheless crucial to assessing the efficacy of CHAMPS. First, it is necessary to establish whether and to what extent students enrolled in the program demonstrate positive change in their musical achievement. Second, it is of interest to examine whether each student's degree of improvement in their achievement is predicted by their level of engagement in the program.

The remainder of this report describes how these questions were addressed, the answers that were reached, and what these answers may reveal about both the program and the students enrolled in it.

Methods

Participants. All students attending University Heights Charter School (UHCS) and enrolled in grades 4 through 7 were invited to participate in the evaluation. This included 50 students enrolled in CHAMPS, as well as 46 of their peers enrolled in the same grades and classrooms. Of these, parents or guardians of 36 students enrolled in CHAMPS agreed to participate in the evaluation (72.0%), while 24 parents or guardians of students in the comparison group did so (52.3%).

Table 1 reports demographic information for both the CHAMPS and comparison groups. There were no significant differences in the distribution of students between the CHAMPS and comparison groups by gender ($p = .264$), grade ($p = .916$), and classroom ($p = .974$).

Table 1.

	CHAMPS ($n = 36$)		Comparison Group ($n = 24$)	
	Count	Percent	Count	Percent
Gender				
- Male	10	27.8	10	41.7
- Female	26	72.2	14	58.3
Grade				
- 4	6	16.7	4	16.7
- 5	15	41.7	8	33.3
- 6	8	22.2	6	25.0
- 7	7	19.4	6	25.0
Classroom				
- Brandeis	5	13.9	3	12.5
- Caldwell	4	11.1	3	12.5
- Georgetown	4	11.1	2	8.3
- Montclair	7	19.4	3	12.5
- Nyack	2	5.6	2	8.3
- Pace	3	8.3	4	16.7
- Ramapo	7	19.4	5	20.8
- Rutgers	4	11.1	2	8.3

Procedures. The CHAMPS program ran for 29 weeks during the 2013-14 academic year, from October 2013 to June 2014. Data were collected from children and their parents or guardians within two weeks of the program beginning and end dates at a pair of events held at UHCS. Data were collected from children's UHCS teachers according to the same schedule, while CHAMPS teachers were asked to complete measures three times over the course of the program year: in late November 2013, early March 2014, and mid-May 2014.

For the October pre-CHAMPS data collection, completed measures were received from 59 of 60 parents or guardians, 58 children, and 58 UHCS teachers. At the May post-CHAMPS data collection, measures were received from 33 parents or guardians, 41 children, and 38 UHCS teachers. Note that four children in the sample left UHCS by the end of the 2013-14 academic year, including three who had been enrolled in CHAMPS. CHAMPS teachers completed measures for 35 students in November 2013, 30 in March 2014 (by which point the three students had left the program), and 29 students in May 2014.

Measures.

Risk. Household risk was assessed by parent responses to a series of items that covered the most-commonly cited risk factors: marital status (single or married/living with partner), highest level of education (assessed on a six-point scale from “grade school” to “advanced degree”), poverty, the number of children in the home, and perceived financial strain. Poverty was indexed by income-to-needs ratio (INR), defined as the family’s income divided by the federal poverty level for a family.⁵ Perceived financial strain was assessed using a subset of six items taken from a measure of acute and chronic financial problems developed by Cutrona.⁶ Following the author’s guidelines, a composite financial strain score was calculated as the sum of item responses.

Resilience. Four sources of family-level resilience were examined. The prevalence of family routines was assessed using a set of 8 items adapted from a previously-published measure⁷ that asked how frequently certain activities occurred in the household (e.g., “We have some time to be together as a family every day.”). A composite family routines score was calculated as the mean frequency of activities. Emotional expressiveness was indexed as the mean response across 20 items of Halberstadt’s measure.⁸ Parent-child interaction was assessed as the mean response on the first nine items of Abidin’s Parenting-Child Dysfunctional Interaction Subscale of the Parenting Stress Index,⁹ reworded so to capture positive aspects of the parent-child relationship. Social support was indexed as mean response to the Cutrona and Russel’s Social Provisions Scale.¹⁰

Program Experience. Students’ experience in the program was assessed using levels of engagement as reported by each student’s primary teacher on a 13-item measure developed for this evaluation (see Appendix A). The measure included items that were positive (e.g., “This student is dedicated to the program”) and negative (e.g., “This student acts like they would rather be somewhere else during class”) in their valence, and teachers were asked to respond on a 5-point scale ranging from “strongly agree” to “strongly disagree”. A mean score of program engagement was calculated after reverse-coding negative-valence items.

Musical Accomplishment. Students’ musical accomplishment was assessed twice during the course of the program year: once over a two-week period in early to mid-December 2013, and again in late April and early May of 2014. In both cases, three-member juries rated students’ playing as they performed a selection of scales and pieces they had worked on in rehearsal. At both assessments, one of the jury panelists was the Artistic Director of CHAMPS (a violinist by training) and the second was a cellist with the NJSO. The remaining panelist was a professional musician (a horn player for the fall assessment and a violinist in the spring). Each assessment of musical accomplishment yielded a score on a scale from 1 to 63.

⁵ For a fuller description of how income-to-needs ratios are calculated, see McLoyd (1998), cited above.

⁶ Cutrona, C. (2000). *Chronic and Acute Financial Problems*. Iowa City: University of Iowa.

⁷ Jensen, E. W., James, S. A., Boyce, W. T., & Hartnett, S. A. (1983). The family routines inventory: Development and validation. *Social Science and Medicine*, 17, 201-211.

⁸ Halberstadt, A. G. (1986). Family socialization of emotional expression and nonverbal communication styles and skills. *Journal of Personality and Social Psychology*, 51, 827-836. The 20 items used to calculate mean response corresponded to those items that loaded on the positive-dominant and positive-submissive factors in Cassidy, J., Parke, R. D., Butkovsky, L., & Braungart, J. M. (1992). Family-peer connections: The roles of emotional expressiveness within the family and children’s understanding of emotions. *Child Development*, 63, 603-618.

⁹ Abidin, R. R. (1983). *Parenting Stress Index – Manual*. Charlottesville, VA: University of Virginia Press.

¹⁰ Cutrona, C. E., & Russell, D. (1987). The provisions of social relationships and adaptation to stress. In W. H. Jones & D. Perlman (Eds.), *Advances in Personal Relationships*. Greenwich, CT: JAI Press.

Perseverance. Perseverance was assessed using child responses to the 12-item grit scale,¹¹ as well as parent and teacher responses to adapted versions of the scale developed for this evaluation. Composite grit scores were calculated for each respondent as the mean of endorsed items.

Prosocial Behaviors. Prosocial behaviors were assessed using Goodman's Strengths and Difficulties Questionnaire,¹² supplemented with data provided by UHCS. Items from four subscales of the child version of the Strengths and Difficulties Questionnaire (conduct problems, hyperactivity, peer problems, and prosocial behaviors) were completed by children, while the parallel items from the adult version were completed by parents and UHCS teachers. Composite scores for children, parents, and teachers were calculated as the mean of endorsed items.

Conduct grades for English and math were provided for each trimester on a three-point scale (outstanding, satisfactory, not satisfactory), along with the number of scholar dollars (earned and lost depending on student behavior) at year's end. The number of absences and tardies over the course of the academic year was also provided for each student.

Intrapersonal Skills. Items assessing intrapersonal skills such as emotional maturity and responsibility were administered to children, parents, and UHCS teachers. These items were taken from the Behavioral and Emotional Rating Scale, 2nd edition (BERS-2).¹³ A composite intrapersonal skills score was calculated as the mean of endorsed items.

Future Orientation. Children and their parents or guardians were asked to respond to items adapted from a previously-published measure of future orientation.¹⁴ These items asked how likely children and parents thought it was that students would achieve major life goals (e.g., having a successful career), some of which were particularly appropriate for students enrolled in middle school (e.g., graduating from high school).

Academic Achievement. Students' academic achievement was assessed using assigned grades for English and math, which were reported for each trimester. Grades for each trimester were reported on a four-point scale (A, B, C, F).

Additional Measures. At the spring data collection a small number of additional measures were piloted for possible inclusion in future years of the study. Academic self-concept was assessed using child responses to the Scholastic Competence subscale from Harter's Self-Perception Profile for Children.¹⁵ Goal orientation was assessed by taking the mean of children's responses to the mastery goal orientation and performance-approach goal orientation subscales of the revised Patterns of Adaptive Learning Scales.¹⁶ Cooperation and empathy were measured using child and parent responses to subscales of the same names taken from the Social Skills Improvement System (SSIS).¹⁷ Composite scores for each combination of respondent and construct were calculated as the mean of endorsed items.

¹¹ Duckworth et al., 2007.

¹² Goodman, R. (2006). *The Strengths and Difficulties Questionnaire*. Available online at: www.sdqinfo.com.

¹³ Epstein, M. H. (2004). *BERS-2: Behavioral and Emotional Rating Scale*. Austin, TX: Pro-Ed.

¹⁴ Longscan. (1998). *Future Events Questionnaire*. Chapel Hill, NC: The Longitudinal Studies of Child Abuse and Neglect.

¹⁵ Harter, S. (2012). *Self-Perception Profile for Children: Manual and Questionnaires*. Denver, CO: University of Denver, Department of Psychology.

¹⁶ Midgley, C., Maehr, M. L., Hruda, L. Z., Anderman, E., Anderman, L., et al. (2000). *Manual for the Patterns of Adaptive Learning Scales*. Ann Arbor, MI: University of Michigan.

¹⁷ Gresham, F., & Elliott, S. N. (2008). *Social Skills Improvement System (SSIS)*. San Antonio, TX: Pearson Assessments.

Data Reduction.

Risk and Resilience. Significant correlations were observed between single-parent status and lower levels of education ($r(49) = -.428, p = .002$), single-parent status and lower income-to-needs ratios ($r(49) = -.355, p = .012$), and higher education and income-to-needs ratios ($r(53) = .525, p < .001$). Single-parent status, an income-to-needs ratio less than 1.00 (corresponding to 100% of the federal poverty level), and a highest level of education of a high school diploma or less were coded as indicative of risk. A cumulative risk score was calculated for each family as the proportion of these three variables scored 'at risk' divided by the number of variables endorsed ($M = .428, SD = .356$).

The frequency of family routines was correlated with higher levels of emotional expressiveness ($r(52) = .565, p < .001$), more robust social support ($r(47) = .378, p = .009$), and more positive parent-child relationships ($r(47) = .443, p = .002$). Higher levels of emotional expressiveness were associated with more social support ($r(48) = .500, p < .001$) and more positive parent-child relationships ($r(49) = .361, p = .011$). Therefore a composite resilience score was calculated for each family as the mean of the standardized scores for family routines, emotional expressiveness, social support, and parent-child relationship.¹⁸

Program Experience. An exploratory factor analysis was performed for teacher-reported levels of engagement collected in November 2013 and March and May of 2014. For the first round of engagement measures, the first eigenvalue was 7.10, while the next was 1.86, with final communality estimates ranging from .626 to .931 across items. For the second round the first two eigenvalues were 6.71 and 2.00 (communality estimates = [.610, .866]) and for the third round eigenvalues were 6.67 and 2.79 (communality estimates = [.709, .920]). These results suggested that all items on the measure loaded onto a single factor representing engagement in the program, and therefore a composite engagement score was calculated for each round as the mean score across items. These mean scores were highly inter-correlated ($r[27, 29] = [.588, .866], p < .001$) and therefore an overall measure of engagement was calculated as the mean of scores across rounds of collection.

Perseverance. At the post-CHAMPS data collection perseverance scores as reported by children and parents were correlated ($r(33) = .497, p = .003$), while teacher-reported scores were not correlated with either child or parent scores. These results recommended against calculating a cross-informant composite perseverance score, but suggested that a Bonferroni correction ($\alpha/2 = .025$) be employed when examining perseverance scores.

Prosocial Behaviors. While parent and teacher responses on the Strengths and Difficulties Questionnaire were correlated ($r(22) = .623, p = .002$), both were uncorrelated with child responses, again suggesting that informants' scores be analyzed independently after invoking a Bonferroni correction ($\alpha/2 = .025$). Third-trimester conduct grades for English and math were correlated ($r(58) = .590, p < .001$), and therefore a composite conduct grade was calculated as the mean of conduct grades for English and math grades for both the first and third trimesters. Year-end scholar dollars were correlated with third-trimester conduct grades in English ($r(55) = .490, p < .001$) and math ($r(55) = .562, p < .001$), indicating that a Bonferroni correction ($\alpha/3 = .017$) should be employed when examining scholar dollars or the composite conduct grade. This correction was also recommended for examinations of school absences and tardies (though here a correction to ($\alpha/2 = .025$ is required), which were also correlated ($r(58) = .367, p = .005$).

Intrapersonal Skills. Parent and teacher intrapersonal skill scores were correlated with each other ($r(21) = .552, p = .009$), but neither was associated with child scores. As with perseverance and

¹⁸ z scores were used for standardized scores, allowing responses across measures to be placed on the same metric.

strengths and difficulties scores, this recommends employing a Bonferroni correction ($\alpha/2 = .025$), but does not support the calculation of a cross-informant index score.

Future Orientation. Parent and child future orientation scores were uncorrelated, indicating that no composite is suggested and no Bonferroni correction is necessary.

Academic Achievement. Third trimester academic (as opposed to conduct) grades in English and math were correlated ($r(58) = .548, p < .001$), and therefore composite scores were calculated for the first and third trimesters.

Results

Preliminary Analyses. Table 2 presents mean scores at the pre- (fall) and post-CHAMPS (spring) assessments.

Table 2.

	Pre-CHAMPS Assessment						Post-CHAMPS Assessment					
	Comparison Group			CHAMPS Group			Comparison Group			CHAMPS Group		
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>
Risk	21	.524	.343	25	.347	.353	---	---	---	---	---	---
Resilience	24	-.108	.769	29	.222	.635	---	---	---	---	---	---
Perseverance (scale range = 1 – 5)												
- Child report	21	3.59	.629	34	3.45	.645	16	3.56	.607	25	3.37	.633
- Parent report	24	3.50	.610	25	3.47	.538	16	3.56	.568	18	3.71	.429
- Teacher report	24	3.35	.783	32	3.17	.590	16	3.54	.688	19	3.77	.773
Prosocial Behaviors (SDQ; range = 0 – 2)												
- Child report	20	1.60	.268	32	1.40	.269	16	1.62	.198	24	1.48	.318
- Parent report	24	1.62	.260	29	1.64	.234	16	1.60	.252	18	1.78	.150
- Teacher report	23	1.42	.448	33	1.55	.365	16	1.59	.396	20	1.52	.405
Prosocial Behaviors (School-based indices)												
Conduct Grades	22	1.00	.673	34	1.28	.525	23	1.04	.620	35	1.11	.543
Scholar Dollars	---	---	---	---	---	---	23	1293	344.7	33	1529	278.7
Absences	---	---	---	---	---	---	23	8.48	5.43	35	5.37	4.89
Tardies	---	---	---	---	---	---	23	15.9	16.6	35	10.1	13.3
Intrapersonal Skills (range = 1 – 3)												
- Child report	18	2.31	.333	31	2.04	.541	16	2.20	.314	25	.217	.576
- Parent report	23	2.34	.477	28	2.41	.439	16	2.18	.510	18	2.51	.310
- Teacher report	24	2.06	.891	33	2.08	.657	15	2.44	.573	20	2.21	.631
Future Orientation (range = 1 – 5)												
- Child report	17	3.72	1.02	30	4.17	.804	16	4.23	.386	24	4.31	.514
- Parent report	23	4.18	.530	30	3.95	1.09	16	4.14	.458	18	4.37	.318
Academic Grades	23	1.85	.818	35	1.83	.747	23	2.02	.730	35	2.01	.712
Music Assessment	---	---	---	27	14.9	8.22	---	---	---	25	31.0	14.0
Pilot Measures												
- Goal Orientation	---	---	---	---	---	---	16	3.84	.644	21	4.17	.590
- Empathy												
- Child report	---	---	---	---	---	---	16	3.25	.459	19	3.67	.458
- Parent report	---	---	---	---	---	---	16	3.41	.430	18	3.56	.505
- Teacher report	---	---	---	---	---	---	15	3.23	.559	20	3.08	.613
- Cooperation												
- Child report	---	---	---	---	---	---	16	3.18	.478	21	3.46	.460
- Parent report	---	---	---	---	---	---	16	3.45	.611	18	3.63	.373
- Teacher report	---	---	---	---	---	---	15	3.37	.602	20	3.21	.692

¹ Data were collected only at the pre-assessment

² Data were reported only at year's end

³ Data were collected only for students enrolled in CHAMPS

⁴ Pilot data were collected only at post-assessment

An inspection of the data reported in Table 2 reveals a number of important points. First, for multiple domains there are substantial differences between scores reported for students in the comparison and CHAMPS groups. Parents of students in the comparison group reported higher levels of risk ($t(44) = 1.72, p = .093$) and lower levels of resilience ($t(51) = -1.71, p = .093$) than parents in the comparison group. At the post-CHAMPS assessment, parents of students enrolled in CHAMPS reported higher levels of prosocial behaviors as indexed by the Strengths and Difficulties Questionnaire (SDQ; $t(32) = 2.57, p = .015$), even though no such difference was observed at the pre-CHAMPS assessment ($p = .781$). Similar results were found for parents' ratings of intrapersonal strengths and future orientation, with parents of CHAMPS students reporting higher intrapersonal skills ($t(32) = 2.33, p = .026$) and future orientation ($t(32) = 1.72, p = .095$) scores at the post-assessment, but not the pre-assessment ($p = .602$ and $.364$, respectively). At year's end, students in CHAMPS were found to have more scholar dollars ($t(54) = 2.82, p = .007$) and fewer absences ($t(56) = 2.27, p = .027$).

It is also evident from table 2 that substantial portions of data were missing at the post-CHAMPS assessment for measures of perseverance, prosocial behaviors (SDQ), intrapersonal skills, and future orientation. Missingness for each of these measures was regressed on child age, gender, UHCS classroom and scores on the same measure at the pre-CHAMPS assessment. Teacher-reported perseverance and intrapersonal skills scores were missing as a function of grade and pre-CHAMPS scores, such that older students were more likely to be missing both perseverance ($B = 2.82, p = .016$) and intrapersonal skills scores ($B = 2.86, p = .022$), and students with lower pre-CHAMPS perseverance ($B = -3.06, p = .036$) and intrapersonal skills scores ($B = -3.49, p = .046$) scores were more likely to be missing post-CHAMPS scores. No other post-CHAMPS scores were systematically missing as a function of age, gender, classroom, or pre-CHAMPS scores. Therefore all post-CHAMPS measures were classified as missing at random.¹⁹

Two other points were addressed in the preliminary analyses. First, it was possible that scores on measures of perseverance, the strengths and difficulties questionnaires (SDQ), conduct grades, intrapersonal skills, future orientation, and academic grades were 'nested' as a function of UHCS classroom. That is, scores for students in the same classroom on any or all of these measures may be more similar than scores for students in different classrooms. While this applies most obviously to teacher-reported measures, it is also possible that child- or parent-report measures would display this same characteristic. To investigate this possibility, a series of models were estimated to partition the variance in each measure into its child- and classroom-related components. In no case was the portion of variance attributable to classroom statistically significant, indicating that data were likely not nested as a function of classroom.

Second, it was also possible that scores on these same measures were not normally distributed (i.e., that they were negatively or positively skewed, with disproportionately large numbers of extremely high or low values). While school absences ($G_1 = .854, SE = .314$) and tardies ($G_1 = 1.05, SE = .314$) exhibited an acceptable degree of positive skew, child- and parent-reported strengths and difficulties and parent-reported intrapersonal skills all evidenced an unacceptable degree of negative skew. Further inspection of the data revealed that in each case skew was attributable to a single outlier (defined as $\pm 3 SD$ from the mean), and that with these values removed data on these measures approximated a normal distribution.

As noted in the methods section, a number of measures were piloted in the spring for possible inclusion in future rounds of the evaluation. With the exception of Harter's measure of achievement motivation, students who completed these pilot measures provided reasonably complete responses

¹⁹ Alison, P. D. (2001). *Missing Data*. Thousand Oaks, CA: Sage.

that exhibited good distributional properties. This indicated that these measures should be included in future years of the evaluation (see Discussion section below).

Model Specification and Testing. With these preliminary analyses completed, we proceeded to specify and test the models necessary to answer the guiding questions of the evaluation. We began by examining whether *enrollment* in CHAMPS during the 2013-2014 academic year was associated with outcomes of interest. Thus for this first set of analyses experience in CHAMPS was defined dichotomously (i.e., 0 = not enrolled, 1 = enrolled) and the analyses were applied to the entire sample, comprising students in CHAMPS as well as those in the comparison group.

In most cases, the effects of enrollment in CHAMPS were assessed with respect to levels on a particular measure at the post-CHAMPS assessment after controlling for (or holding constant) scores at the pre-CHAMPS assessment.²⁰ However, in those cases where pre-CHAMPS assessment data were not available (e.g., school absences), the effects of enrollment were necessarily assessed only with respect to year-end data. Unless otherwise noted, the effects of enrollment in CHAMPS were assessed after controlling for gender, grade, and classroom. The possibility that the effects of enrollment varied as a function of students' gender was investigated by including the interaction of enrollment and gender as the last step in each model.²¹

Guiding Questions, Set 1: Direct and Mediating Effects

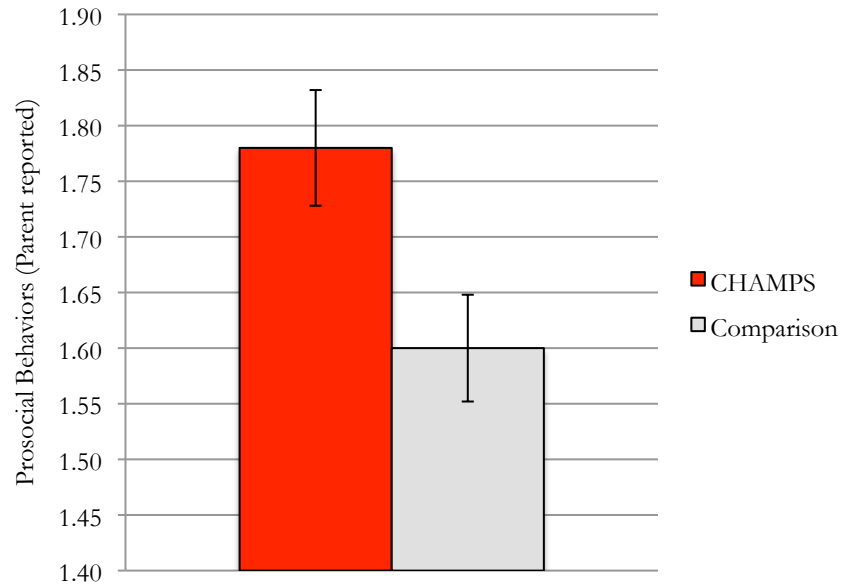
There was no evidence that enrollment in CHAMPS predicted higher levels of academic achievement in the third trimester after controlling for achievement in the first trimester ($p = .933$), and there was also no evidence that enrollment in CHAMPS predicted higher levels of perseverance as reported by children ($p = .583$), parents ($p = .177$), or UHCS teachers ($p = .666$). However, enrollment in CHAMPS was associated with higher parent ratings of children's prosocial behaviors on the Strengths and Difficulties Questionnaire ($p = .011$; see Figure 3).²² No differences were observed for child ($p = .388$) or teacher ($p = .107$) ratings of strengths.

²⁰ Models that included the pre-CHAMPS assessment data may be said to yield residualized change scores. However, throughout the Results and Discussion section, we refer only to whether scores at the post-CHAMPS assessment were higher or lower as a function of enrollment and/or engagement in CHAMPS, as a true consideration of *change* over time would require most than two data points. As noted in the Discussion section, the principal advantage of including pre-CHAMPS assessment data in our analyses was to control for between-group or inter-individual differences in scores prior to program implementation.

²¹ All models were estimated using full information maximum likelihood (FIML) to account for missing data on the dependent variables. Given that data were missing at random, FIML was an acceptable strategy provided that all independent variables related to rates of missingness (grade and pre-CHAMPS scores in the cases of teacher-rated perseverance and intrapersonal skills) were included in the model. Although random effects ANOVAs did not indicate nesting in any outcomes as a function of classroom, it was possible that non-significant estimates of variance associated with classroom were a function of small sample size. Therefore all models specified that students were nested within classrooms. Degrees of freedom were estimated according to the Kenwood Rogers method, which is most appropriate for small sample sizes.

²² As noted above, parent and teacher responses on the post-CHAMPS administration of the Strengths and Difficulties Questionnaire were positively correlated, necessitating a Bonferroni correction when assessing the statistical significance of main effects. The main effect of enrollment on parent ratings was robust to this correction ($p = .011 < \alpha = .05/2 = .025$).

Figure 3. Differences in Parent-Reported Prosocial Behaviors by Enrollment in CHAMPS



Other measures of prosocial behavior were available only at year's end. Initial models indicated that enrollment in CHAMPS was associated with fewer school absences ($p = .020$), but while this effect was robust to Bonferroni correction, it was not robust to risk (i.e., with risk included in the model the effect of enrollment was no longer significant).²³ Enrollment was associated with higher numbers of scholar dollars ($p = .057$) and fewer tardies ($p = .097$) at levels approaching significance.

Enrollment in CHAMPS was associated with higher child ratings of their own intrapersonal skills for girls ($p = .027$; see Figure 4), as well as higher parent ratings of intrapersonal skills for students of both genders at a level that approached significance ($p = .064$). Enrollment was also associated with higher child ratings of future orientation among girls, but not among boys ($p = .011$; see Figure 5). Table 3 summarizes the effects of enrollment on outcomes.

²³ The Bonferroni correction is necessitated by the positive correlation between absences and tardies, while the inclusion of risk was necessitated because risk predicted absences (see below).

Figure 4. Differences in Child-Reported Intrapersonal Skills by Enrollment in CHAMPS

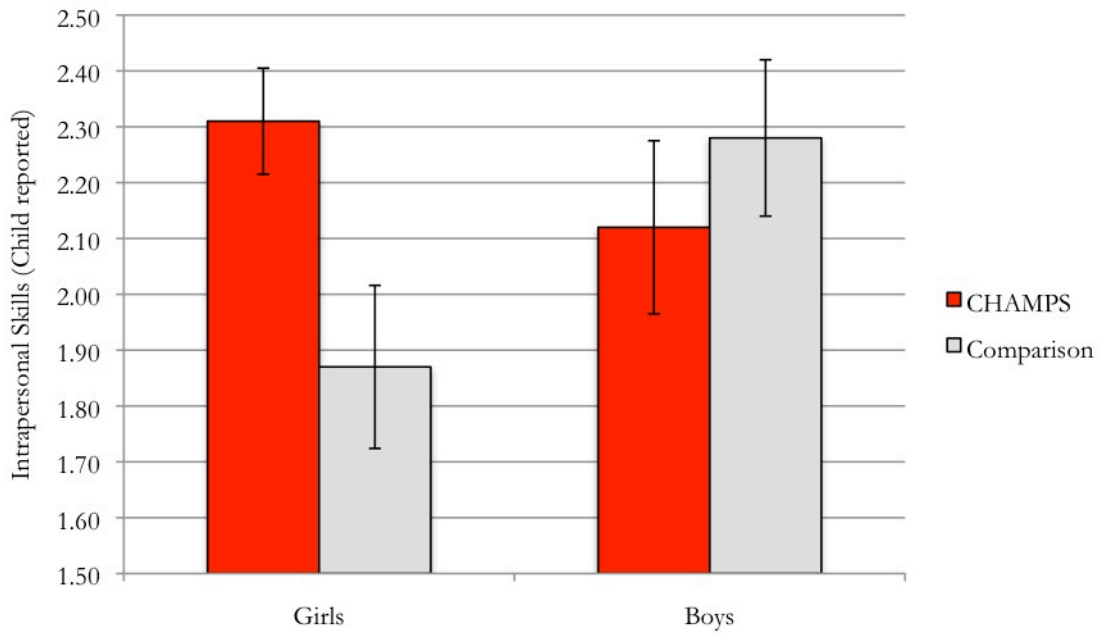


Figure 5. Differences in Child-Reported Future Orientation by Enrollment in CHAMPS

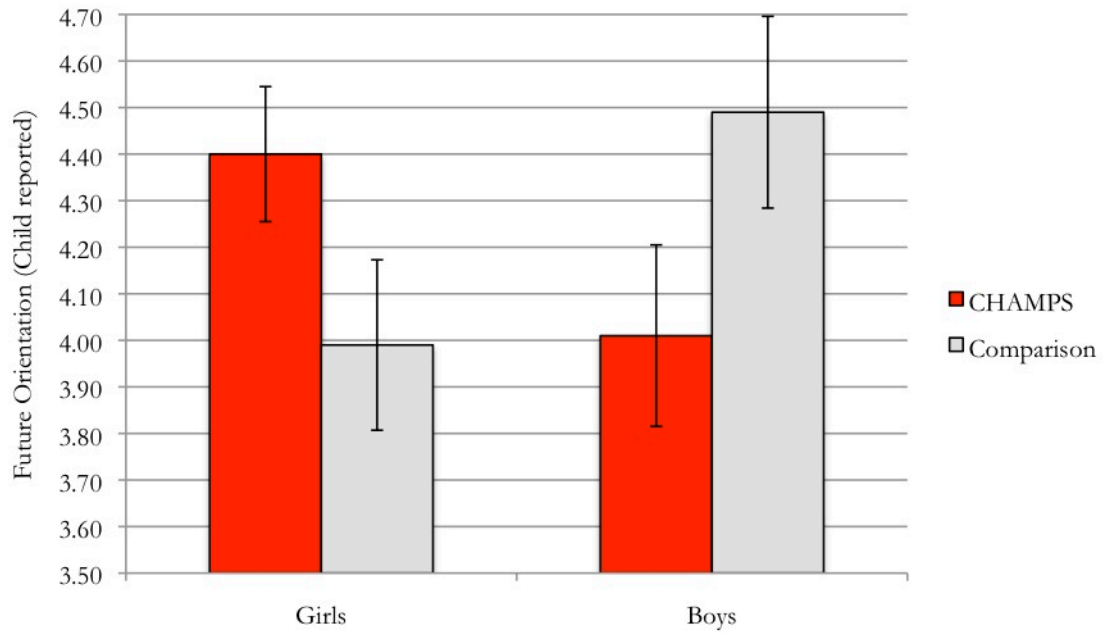


Table 3

Domain	Respondent	Effect Type		Effect Size				
				Least-Squares Means			SD ²	Effect
				CHAMPS (SE) ¹	Comparison (SE)	Difference		
Prosocial Behaviors (SDQ)	Parent	Main		1.78 (.052)	1.60 (.048)	0.18	0.24	0.75
Intrapersonal Strengths (BERS)	Child	Interaction	Girls	2.31 (.095)	1.87 (.146)	0.44	0.48	0.92
			Boys	2.12 (.155)	2.28 (.140)	-0.16	0.47	0.34
Future Orientation	Child	Interaction	Girls	4.40 (.145)	3.99 (.183)	0.41	0.46	0.89
			Boys	4.01 (.195)	4.49 (.206)	-0.48	0.48	1.00

¹ Standard error of the least-squares means estimate.

² For interaction effects, the standard deviation for each gender was used in calculating effect sizes.

Next we examined whether a student's level of engagement in CHAMPS – defined by the composite measure of engagement described above – would predict differences in academic achievement or more proximal outcomes. These analyses were necessarily confined to students enrolled in CHAMPS during the 2013-2014 academic year.

While there was no evidence that engagement in CHAMPS was associated with higher third trimester academic achievement after controlling for first trimester achievement ($p = .359$), higher levels of engagement in CHAMPS were associated with higher ratings of perseverance. More highly-engaged students of both genders were rated as exhibiting higher levels of perseverance by UHCS teachers ($p = .001$; see Figure 6),²⁴ while parents of more highly-engaged boys also rated their children as displaying higher levels of perseverance ($p = .016$).

Significant effects of engagement were not observed for parent ratings of strengths as assessed by the Strengths and Difficulties Questionnaire ($p = .711$), but they were observed for child and teacher ratings. Higher levels of engagement were associated with higher child ratings of their own prosocial behaviors among boys ($p = .015$; see Figure 7) and higher levels of teacher-reported prosocial behaviors among students of both genders ($p = .006$; see Figure 8). Among the measures of prosocial behavior provided by the school, higher levels of engagement corresponded with more scholar dollars ($p = .040$) after controlling for risk.

Levels of engagement were not associated with changes in child ($p = .260$), parent ($p = .100$), or teacher reports of intrapersonal strengths ($p = .850$), nor were they associated with child ($p = .517$) or parent reports of changes in future orientation ($p = .458$). Table 4 summarizes the significant effects of level of engagement.

²⁴ The effects of level of engagement on parent and teacher perseverance scores were robust to the Bonferroni correction necessitated by the positive correlation between these scores. For purposes of estimating effect sizes, low engagement was defined as $-1 SD = 3.37$ and high engagement was defined as $+1 SD = 4.80$. In the case of parent-reported perseverance, the estimate of the least-squared mean fell outside of the range of observed scores for highly-engaged students, indicating that the extremely large effect size must be interpreted with caution.

Figure 6. Differences in Teacher-Reported Perseverance by Engagement in CHAMPS

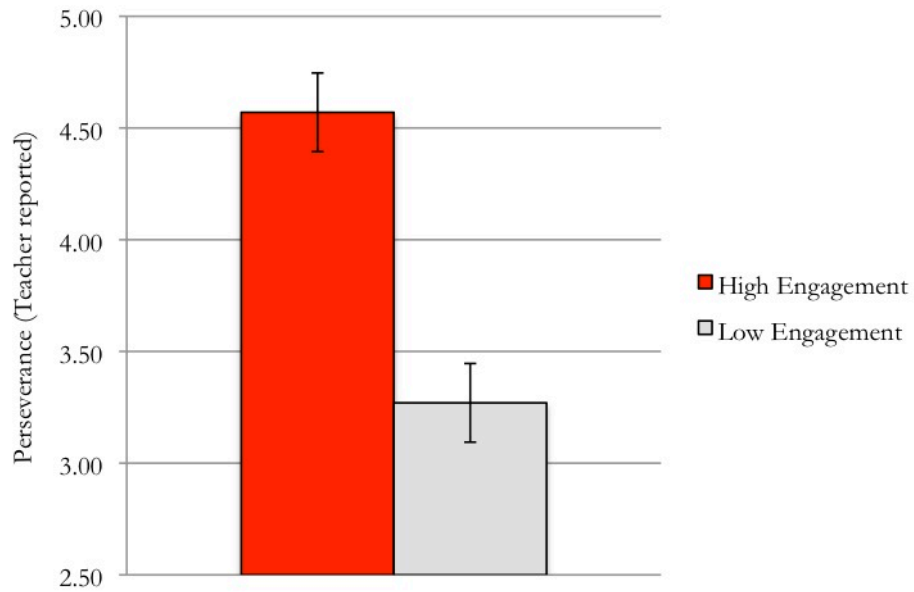


Figure 7. Differences in Child-Reported Prosocial Behaviors by Engagement in CHAMPS

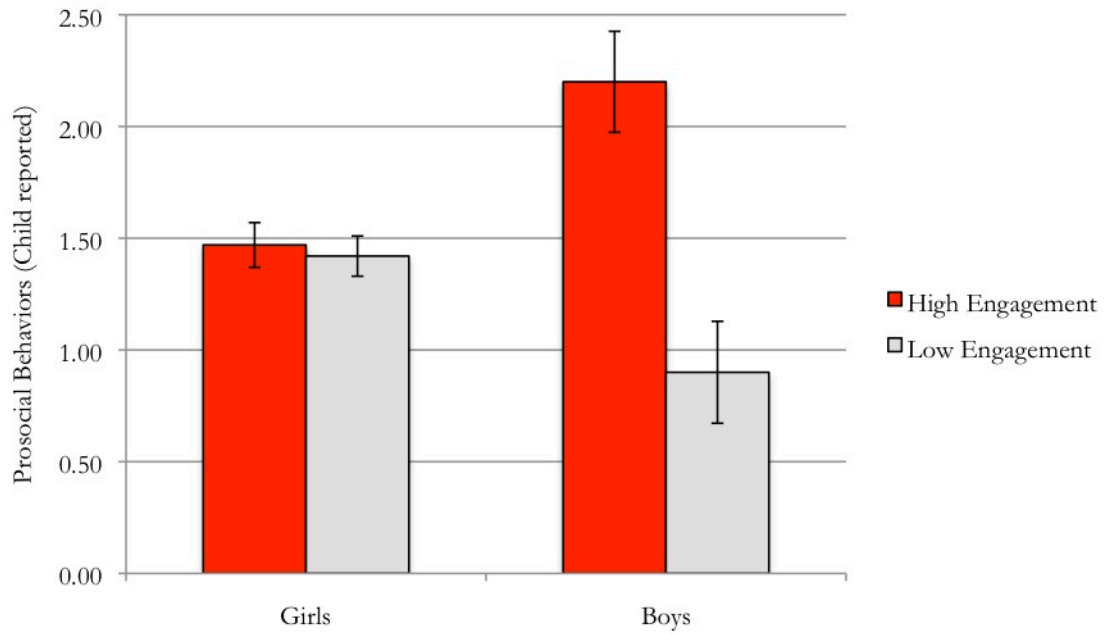


Figure 8. Differences in Teacher-Reported Prosocial Behaviors by Engagement in CHAMPS

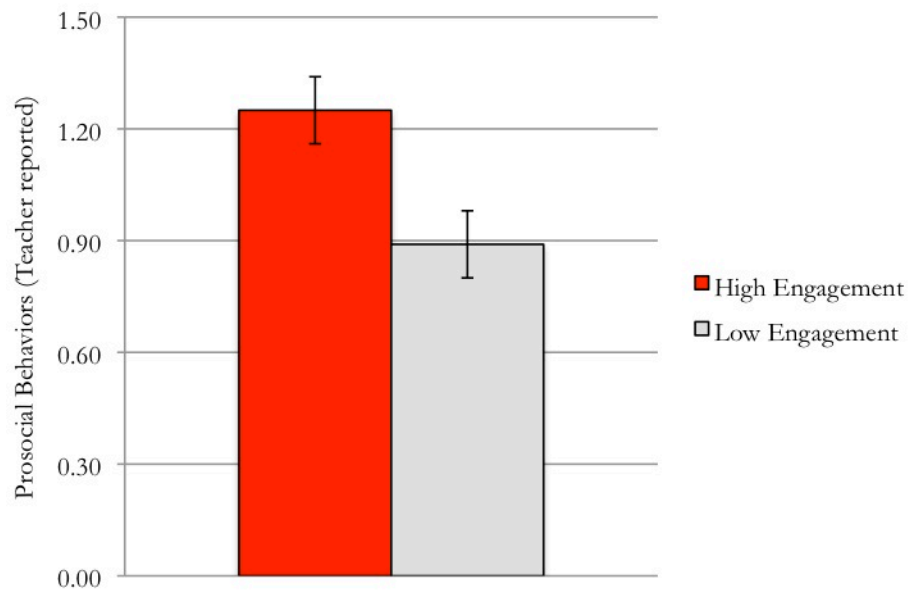


Table 4

Domain	Respondent	Effect Type		Effect Size				
				Least-Squares Means ³			SD ⁴	Effect
				High Eng. (SE) ²	Low Eng. (SE)	Difference		
Perseverance	Parent	Interaction ¹	Girls	3.88 (.110)	3.71 (.180)	0.17	0.39	0.44
			Boys	7.70 (.982)	.03 (1.19)	7.67	0.52	14.8
	Teacher	Main		4.57 (.176)	3.27 (.176)	1.30	0.77	1.69
Prosocial Behaviors (SDQ)	Child	Interaction ¹	Girls	1.47 (.100)	1.42 (.090)	0.05	0.27	0.19
			Boys	2.20 (.226)	0.90 (.228)	1.30	0.42	3.10
	Teacher	Main		1.25 (.090)	0.89 (.090)	0.36	0.41	0.88
Scholar dollars	n/a	Main		1606 (81.8)	1388 (81.1)	218	279	0.78

¹ Ratings of engagement were mean-centered prior to being included in interaction terms. Interactions were probed using low (-1 *SD* = -.71) and high (+1 *SD* = .71) levels of engagement.

² Standard error of the least-squares means or model-implied estimate (see note 3).

³ Note that for main effects the model-implied estimates at each level of engagement are reported, rather than least squares means.

⁴ For interaction effects, the standard deviation for each gender was used in calculating effect sizes.

The remaining item from the first set of guiding questions concerned mediation, and asked whether the effects of program experience on academic achievement are mediated by its effects on more proximal outcomes. However, no effects of either program enrollment or engagement on academic achievement were observed, thus precluding the possibility of mediation.

Guiding Questions, Set 2: Indirect or Mitigating Effects

The second set of guiding questions examined the possibility that either enrollment or engagement in CHAMPS might exert an indirect effect on academic achievement or proximal outcomes by mitigating or buffering the effects of risk. Thus the first question of this set sought to establish whether risk did in fact exert deleterious effects on various outcomes within the context of the study sample. However, neither risk (defined as described above) nor poverty (defined as income-to-needs ratio) predicted levels of academic achievement or any other measure, with the exception of school absences. Without an effect of risk, there is nothing for enrollment or engagement in CHAMPS to mitigate, nor was there a possibility that mitigation effects might be amplified by other sources of resilience in the child's home.

However, these facts did not exclude the possibility that enrollment may amplify the effects of these sources of resilience on academic achievement or other proximal processes. To investigate this possibility, a new set of models were run in which post-CHAMPS scores for each outcome were modeled as a function of gender, grade, household resilience, enrollment in CHAMPS, and, where available, pre-CHAMPS scores. However, only in one instance did the interaction of resilience and enrollment predict an outcome. At low levels of resilience (defined as -1 *SD* on the composite resilience α -score), there was no appreciable difference in scholar dollars between students enrolled in the program (who ended the year with 1311 dollars, on average) and those not enrolled (1362). However, at high levels of resilience, students enrolled in CHAMPS ended the year with significantly more scholar dollars (1602) than their peers (1405).

The final set of guiding questions concerned students' musical progress, and to what extent this progress was predicted by each student's level of engagement in the program. As can be seen above in table 2, the mean score on the post-CHAMPS assessment of musical achievement was nearly twice that observed at the pre-CHAMPS assessment. On average, students gained 14.6 (*SD* = 15.1) points on a measure that included 63 possible points, corresponding to a statistically-significant change in scores ($p < .001$). Gains ranged from 7 to 44 points, with nearly three-quarters (72.7%) of students exhibiting positive change in scores. Higher levels of engagement were associated with higher post-CHAMPS scores after controlling for pre-CHAMPS scores at a rate approaching significance ($p = .099$). This finding was robust to student grade, gender, and program teacher.²⁵

²⁵ Although the estimate of variance in post-CHAMPS musical accomplishment attributable to program teacher was not statistically significant, the possibility that accomplishment scores were nested by teacher was accommodated in the model.

Discussion

Summary of Findings. Preliminary analyses revealed that parents of students enrolled in CHAMPS reported higher levels of prosocial behaviors (as indexed by the SDQ), intrapersonal skills, and future orientation following participation in the program (i.e., at the post-CHAMPS assessment), despite the fact that these differences were not observed prior to participation. Moreover, students enrolled in CHAMPS were found to have more scholar dollars and fewer absences over the course of the academic year. These findings may be taken as evidence that participation in CHAMPS is associated with better outcomes for students, but they constitute a fragile form of evidence for two reasons: first, students were not assigned at random to participate in CHAMPS, but rather self-selected into the program. Thus the same factors that contributed to students' (or parents') choice to be in the program may account for the apparent effects of the program. Second, other factors – including gender, grade, classroom, risk and resilience – may account for the differences observed at year's end. Stronger forms of evidence are provided by the series of models presented above, which control for students' initial levels on measures wherever pre-CHAMPS assessment data were available.

There was no evidence that participation in CHAMPS was associated with higher levels of academic achievement, regardless of whether participation was defined as enrollment in the program or as students' level of engagement. Why might this be so? Academic achievement is difficult to influence – many programs designed explicitly for that purpose fail to do so. It is also an outcome that is affected by many contextual factors beyond those measured in this evaluation. It is therefore fair to ask whether a program intended, first and foremost, to teach music can be asked to influence academic achievement, particularly in the course of a single year. Indeed, it is possible that the effects of CHAMPS on academic achievement would take years to become manifest, a possibility that should not be discounted.

However, participation in CHAMPS – defined both as enrollment and engagement – was associated with higher scores on a number of more proximal outcomes. Enrollment in CHAMPS was associated with higher levels of prosocial behaviors as reported by parents, and the effect size of CHAMPS was 0.75. Effect sizes provide contextual information about the strength or magnitude of an effect by putting all such effects on a common scale, ranging from 0 to infinity (in theory, if not in practice). When comparing differences between groups (as it being done here), an effect size of 0.80 or greater is considered large, an effect between 0.40 and 0.70 medium or moderate, and an effect between 0.10 and 0.30 small or modest.²⁶ Thus, in the case of parent ratings of prosocial behaviors, enrollment in CHAMPS was associated with a large effect size.

Enrollment in CHAMPS was also associated with higher child ratings of their own of intrapersonal skills and future orientation, but only for girls. The effect of enrollment on both of these domains was large: 0.92 for intrapersonal skills and 0.89 for future orientation. For boys, the association between enrollment in CHAMPS and ratings of one's own intrapersonal skills and future orientation were in the opposite direction. Year-end ratings of intrapersonal skills and future orientation were *lower* among boys enrolled in CHAMPS than among their comparison-group peers after controlling for initial levels of both measures. In the case of intrapersonal skills, the small size of the effect (0.34) indicates that this difference may not be of concern, but for future orientation the effect size is large (1.00) and therefore must be addressed.

²⁶ Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Mahwah, NJ: Erlbaum.

It has long been known that while children of both genders experience declines in self-esteem during adolescence, this decline is far more precipitous for girls.²⁷ Moreover, boys tend to overestimate their abilities at this age, while girls underestimate them.²⁸ Thus it may be the case that for both girls and boys, enrollment in CHAMPS exerts a recalibrating effect where children's expectations for the future are concerned. The experience of working with talented teaching artists and learning about the broader world of music and professional life may correct girls' misconceptions about their abilities and prospects for the future, whereas for boys these same experiences may serve as a reality check about the amount of work that professional success demands. From a practical standpoint of program design, the question becomes how to maintain this positive recalibration for girls while reducing the negative recalibration for boys.

In addition to the effects of enrollment, there was also evidence that students' levels of engagement in CHAMPS were associated with differences in proximal outcomes. Students who were more highly engaged in CHAMPS were rated by their teachers as being more persistent and as exhibiting higher levels of prosocial behaviors. Note that these results take into account teachers' ratings of perseverance and prosocial behaviors prior to CHAMPS, so it is not merely the case that students rated as higher by teachers on these measures in the fall are also being rated more highly in the spring. Instead, these results indicate that after controlling for students' initial scores, their scores in the spring were also higher. This, combined with the robust effect sizes for teachers' ratings of perseverance (1.69) and prosocial behaviors (0.88) suggests that students who were more highly engaged in CHAMPS realized more substantial benefits to these domains.

Higher levels of engagement in CHAMPS were also associated with benefits to other proximate domains, though these benefits were realized disproportionately by boys. More highly-engaged children were rated as being more perseverant by their parents; for girls, the size of this effect was moderate (0.44), but for boys it was very large (14.8). Similarly, children's ratings of their own prosocial behaviors were higher for more highly engaged children, but whereas for girls the size of this effect was modest (0.19), for boys it was large (3.10).

These results, taken together with those reported for enrollment, suggest that for certain proximate outcome domains there may be both threshold and dosage effects for CHAMPS that vary as a function of gender. For girls, merely being enrolled in CHAMPS is associated with benefits to intrapersonal strengths and future orientation, while for boys the most substantial benefits of the program were realized among those who were most highly engaged. An implication for program design would be to determine how to increase levels of engagement among boys enrolled in the program.

One of the guiding questions for the evaluation was whether participation in the program – defined either as enrollment or engagement in CHAMPS – would buffer the effects of risk on academic achievement or more proximal outcomes. A second question was whether these buffering effects might be amplified by other sources of resilience in the child's home. However, in order for these buffering effects to be observed, it first had to be established that risk exerted a deleterious influence on these outcomes, and this influence was not observed, either in the sample as a whole (including all students in both the CHAMPS and comparison groups) or among students enrolled in CHAMPS.²⁹ This may be due to the limited variability in levels of risk in our sample: all but 13 families were

²⁷ American Association of University Women. (1991). *Shortchanging Girls, Shortchanging America*, Washington DC: Author.

²⁸ Cole et al. (1999). Children's over- and underestimation of academic competence: A longitudinal study of gender differences, depression, and anxiety. *Child Development*, 70, pp. 459 – 473.

²⁹ The exceptions were year-end absences in the full sample and year-end scholar dollars in the sub-sample of students enrolled in CHAMPS. These effects were not buffered by enrollment or engagement, respectively.

classified as at-risk on the basis of either income-to-needs ratio, education, and marital status, and approximately equals numbers of families were classified at-risk on one, two, or three of these factors. The levels of risk observed in our sample were high in absolute terms: over 40% of the sample was below the federal poverty level (\$23,550 for a family of four), and the majority of parents in the sample were single. With these high mean levels, risk may exceed some threshold required to exert a strong but consistent effect across values observed in the sample, even if the values themselves are variable. A third possibility is that we simply lacked sufficient statistical power to detect the effects of risk on either academic or more proximate outcomes.

Limitations and Recommendations. The major limitation of the evaluation was imposed by sample size. As noted above, approximately two-thirds of students enrolled in CHAMPS agreed to participate in the study, while slightly more than half of eligible students joined the comparison group. But this total sample size was reduced to a lesser functional sample size by missing data, particularly for the post-CHAMPS assessment. The more complex models required to provide stronger forms of evidence demand more statistical power than less involved alternatives, and statistical power is influenced strongly by sample size. This may explain why nearly all statistically-significant results were associated with large effect sizes – smaller effect sizes may not have been discernable given the size of our functional sample.

We would recommend that this limitation be addressed in two ways: first, complete or very nearly-complete data must be collected from all consenting participants. Second, efforts should be made to increase the proportion of eligible comparison-group families consenting to participate in the study. Both of these strategies may require larger incentives be offered to families, but the marginal cost of this investment considered against the total cost of the program would be outweighed by the potential benefit.

Conclusion. The CHAMPS program is recently celebrated its first birthday. In its short existence, it has assembled an impressive record of accomplishments. At the most fundamental level, it has offered an intensive, high-quality music education to students who otherwise would not have received it, and in so doing, demonstrably advanced the musical achievement of every three out of four of its enrollees. In the rush to demonstrate other outcomes for the program this fact should not be overlooked. But there is evidence that CHAMPS may contribute to the development of outcomes beyond music, including perseverance, prosocial behaviors, intrapersonal skills, and future orientation. As students continue in the program, and as the program continues to grow with them, the changes in these proximal outcomes may translate into changes in academic performance. But even if they do not, fostering these characteristics in students can have long-lasting consequences for life outcomes that are in the end more important than grades or test scores, and achieving those long-lasting effects is the ultimate goal of CHAMPS: to harness the power of music to change lives.

Works Cited

- Abidin, R. R. (1983). *Parenting Stress Index – Manual*. Charlottesville, VA: University of Virginia Press.
- Alison, P. D. (2001). *Missing Data*. Thousand Oaks, CA: Sage.
- American Association of University Women. (1991). *Shortchanging Girls, Shortchanging America*, Washington DC: Author.
- Brooks-Gunn, J. & Duncan, G. J. (1997). The effects of poverty on children. *The Future of Children*, 7, 55-71.
- Cassidy, J., Parke, R. D., Butkovsky, L., & Braungart, J. M. (1992). Family-peer connections: The roles of emotional expressiveness within the family and children's understanding of emotions. *Child Development*, 63, 603-618.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Mahwah, NJ: Erlbaum.
- Cole et al. (1999). Children's over- and underestimation of academic competence: A longitudinal study of gender differences, depression, and anxiety. *Child Development*, 70, pp. 459 – 473.
- Cutrona, C. (2000). *Chronic and Acute Financial Problems*. Iowa City: University of Iowa.
- Cutrona, C. E., & Russell, D. (1987). The provisions of social relationships and adaptation to stress. In W. H. Jones & D. Perlman (Eds.), *Advances in Personal Relationships*. Greenwich, CT: JAI Press.
- Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology*, 92, 1087-1101.
- Epstein, M. H. (2004). *BERS-2: Behavioral and Emotional Rating Scale*. Austin, TX: Pro-Ed.
- Goodman, R. (2006). *The Strengths and Difficulties Questionnaire*. Available online at: www.sdqinfo.com.
- Gresham, F., & Elliott, S. N. (2008). *Social Skills Improvement System (SSIS)*. San Antonio, TX: Pearson Assessments.
- Halberstadt, A. G. (1986). Family socialization of emotional expression and nonverbal communication styles and skills. *Journal of Personality and Social Psychology*, 51, 827-836.
- Harter, S. (2012). *Self-Perception Profile for Children: Manual and Questionnaires*. Denver, CO: University of Denver, Department of Psychology.
- Jensen, E. W., James, S. A., Boyce, W. T., & Hartnett, S. A. (1983). The family routines inventory: Development and validation. *Social Science and Medicine*, 17, 201-211.
- Longscan. (1998). *Future Events Questionnaire*. Chapel Hill, NC: The Longitudinal Studies of Child Abuse and Neglect.
- McLoyd, V. C. (1998). Socioeconomic disadvantage and child development. *American Psychologist*, 53, 185-204.

- Midgley, C., Maehr, M. L., Hruda, L. Z., Anderman, E., Anderman, L., Freeman, K. E., Gheen, M., Kaplan, A., Kumar, R., Middleton, M. J., Nelson, J., Roeser, R., & Urdan, T. (2000). *Manual for the Patterns of Adaptive Learning Scales*. Ann Arbor, MI: University of Michigan
- Winner, E., Goldstein, T., & Vincent-Lancrin, S. (2013). *Art for Art's Sake? The Impact of Arts Education*. Washington, DC: Organization for Economic Cooperation and Development.

Appendix A

<p>The following survey asks questions about students in the NJSO pilot program. Please note that although we ask you to indicate both your name and the name of your student, we will not share your responses with anyone.</p>					
Your Name:					
Students Name (first name, last initial):					
How long have you worked with this student (in months)?					
Using the scale below, please indicate how well do you feel you know this student?					
<input type="checkbox"/> Very well	<input type="checkbox"/> Well	<input type="checkbox"/> Not well	<input type="checkbox"/> Hardly at all		
<p>In this section you will be asked how strongly you agree or disagree with each statement about this student. Please answer using your impressions of this student throughout the past program year.</p>					
SA = strongly agree	A = agree	NS = not sure	D = disagree	SD = strongly disagree	
1. This student often forgets his or her instrument or music.	SA	A	NS	D	SD
2. This student has made real progress towards mastery of his or her instrument this year.	SA	A	NS	D	SD
3. This student is dedicated to the program.	SA	A	NS	D	SD
4. This student clearly enjoys class.	SA	A	NS	D	SD
5. This student volunteers to answer questions asked in class.	SA	A	NS	D	SD
6. This student takes care of his or her instrument.	SA	A	NS	D	SD
7. This student pays attention during class.	SA	A	NS	D	SD
8. This student is disrespectful and/disruptive during class.	SA	A	NS	D	SD
9. This student is excited about the opportunity to perform in public.	SA	A	NS	D	SD
10. This student acts like they would rather be somewhere else during class.	SA	A	NS	D	SD
11. This student accepts criticism and works to improve his or her technique based upon it.	SA	A	NS	D	SD
12. This student will help others in a constructive way during class.	SA	A	NS	D	SD
13. This student plays like a member of an ensemble during class by listening to others.	SA	A	NS	D	SD

Thank you for taking the time to complete this survey.